

わかりやすいパターン認識

第6章 EMアルゴリズム

小野寺喜行

教師なし学習に伴う問題

- ・教師なしの場合

$$P(\mathbf{x}) = \int_{\mathbf{s}} P(\mathbf{x}, \mathbf{s}) = \int_{\mathbf{s}} P(\mathbf{s}) P(\mathbf{x} | \mathbf{s})$$

\mathbf{x} についての情報は得られる

\mathbf{s} についての情報は得られない(隠れ変数/潜在変数)

対数尤度 $\log P(\mathbf{x}) = \log \int_{\mathbf{s}} P(\mathbf{x}, \mathbf{s})$



最大化計算は一般に困難

log-sumからsum-logへ

- ・パラメータベクトル $\theta = (\theta_1, \dots, \theta_c, \pi_1, \dots, \pi_c)$ を表記に含める

$$\log P(\mathbf{x}) = \log \underset{\mathbf{s}}{\mathring{A}} P(\mathbf{x}, \mathbf{s}) \longrightarrow \log P(\mathbf{x}; q) = \log \underset{\mathbf{s}}{\mathring{A}} P(\mathbf{x}, \mathbf{s}; q)$$

パラメータの最尤推定値は、上式を最大化する θ に相当

- ・log-sumの形では微分形が有理式で扱いが難しい



イエンゼンの不等式によりsum-logの形の下限値を導出、
その下限値の最大化を図る

log-sumからsum-logへ

- ・次式の目的関数を考える

$$J(q) = \log \prod_{i=1}^c f_i(q)$$

最適化すべきパラメータ θ をもつ任意の関数(正值)



$$\frac{\nabla J(q)}{\nabla q} = \frac{\nabla f_i(q)}{\prod_{i=1}^c f_i(q)} = 0$$

θ に関して微分して0とおく。
通常、解析的に解けない。

log-sumからsum-logへ

- ・ θ^0, θ をそれぞれ更新前、更新後のパラメータとおき

$$h_i^0 = \frac{f_i(q^0)}{\mathop{\text{a}}_{j=1}^c f_j(q^0)} \text{ を導入。} (h_i^0 > 0, \mathop{\text{a}}_{i=1}^c h_i^0 = 1)$$

- ・対数関数が上に凸より、イエンゼンの不等式を用い、 $J(\theta)$ の下限值を導出する。

$$J(q) = \log \mathop{\text{a}}_{i=1}^c f_i(q) = \log \mathop{\text{a}}_{i=1}^c \frac{h_i^0 f_i(q)}{h_i^0} \geq \mathop{\text{a}}_{i=1}^c h_i^0 \log \frac{f_i(q)}{h_i^0}$$

log-sumからsum-logへ

- ・不等式の両辺の差を計算すると

$$\log \prod_{i=1}^c f_i(q) - \prod_{i=1}^c h_i^0 \log \frac{f_i(q)}{h_i^0} = \prod_{j=1}^c h_j^0 \left(\log \prod_{i=1}^c f_i(q) - \log \frac{f_i(q)}{h_j^0} \right)$$

$$= \prod_{j=1}^c h_j^0 \log \frac{h_j^0}{h_j}$$

ただし $h_i = \frac{f_i(q)}{\prod_{j=1}^c f_j(q)}$

X を離散的な値 x_1, x_2, \dots, x_c をとる確率変数、

$h_i^0 = H^0(x_i), h_i = H(x_i)$ ($i = 1, 2, \dots, c$)とすると、

$H^0(X), H(X)$ のKLダイバージェンスに相当。

log-sumからsum-logへ

$$\text{KL}(H^0, H) = \sum_{j=1}^c h_j^0 \log \frac{h_j^0}{h_j}$$



$$J(q) = \sum_{i=1}^c h_i^0 \log \frac{f_i(q)}{h_i^0} + \text{KL}(H^0, H)$$

log-sumからsum-logへ

- ・ θ の代わりに θ^0 を代入

$$J(q^0) = \mathop{\mathring{a}}_{i=1}^c h_i^0 \log \frac{f_i(q^0)}{h_i^0} + \text{KL}(H^0, H^0)$$

$$J(q) - J(q^0) = \mathop{\mathring{a}}_{i=1}^c h_i^0 \log f_i(q) - \mathop{\mathring{a}}_{i=1}^c h_i^0 \log f_i(q^0) + \text{KL}(H^0, H)$$

- ・ $Q(q^0, q) = \mathop{\mathring{a}}_{i=1}^c h_i^0 \log f_i(q)$ (Q関数)とおく

$$J(q) - J(q^0) = Q(q^0, q) - Q(q^0, q^0) + \text{KL}(H^0, H)$$

log-sumからsum-logへ

- ・得られた θ を θ^0 として繰り返し更新

→ $J(\theta)$ は極大点に収束

$$\begin{aligned} \cdot J(q) &\stackrel{c}{=} \sum_{i=1}^c h_i^0 \log \frac{f_i(q)}{h_i^0} = \sum_{i=1}^c h_i^0 \log f_i(q) - \sum_{i=1}^c h_i^0 \log h_i^0 \\ &= Q(q^0, q) - \sum_{i=1}^c h_i^0 \log h_i^0 \end{aligned}$$

θ に関して $Q(\theta^0, \theta)$ を最大化

→ $J(\theta)$ の下限值を最大化することと等価

EMアルゴリズム

Step1. θ^0 の初期値を与える。

Step2. θ^0 を用いて下式より $Q(\theta^0, \theta)$ を求める。

$$Q(q^0, q) = \sum_{i=1}^c h_i^0 \log f_i(q)$$

Step3. θ^0 を定数とみなし、 $Q(\theta^0, \theta)$ を最大にする θ を求める。

Step4. 下式より対数尤度を求め、その増分が閾値以下なら終了。それ以外は $\theta^0 \leftarrow \theta$ とし、Step2に戻り繰り返す。

$$\log P(\mathbf{x}; q) = \log \sum_{\mathbf{s}} P(\mathbf{x}, \mathbf{s}; q)$$

Q関数の特性

- ・ $\log P(\mathbf{x}; q) = \log \mathop{\text{arg}}_s P(\mathbf{x}, \mathbf{s}; q)$ を最大化するためのQ関数

$$Q(q^0, q) = \mathop{\text{arg}}_s P(\mathbf{s} | \mathbf{x}; q^0) \log P(\mathbf{x}, \mathbf{s}; q)$$

- ・ x_t, s_t が各 t に関して統計的に独立とする

$$\begin{aligned} Q(q^0, q) &= \mathop{\text{arg}}_s P(s_1 | x_1; q^0) \cdots P(s_n | x_n; q^0) \\ &\quad \times (\log P(x_1, s_1; q) + \cdots + \log P(x_n, s_n; q)) \\ &= \mathop{\text{arg}}_{t=1}^n Q_t(q^0, q) \end{aligned}$$

Q関数の特性

$$\begin{aligned}
 Q_t(q^0, q) &= \mathop{\text{E}}_{\mathbf{s}} \left[\prod_{i=1}^n P(s_i | x_i; q^0) \times \log P(x_t, s_t : q) \right] \\
 &= \mathop{\text{E}}_{s_t} \left[\mathop{\text{E}}_{\mathbf{s}_{-t}} \left[\prod_{i=1}^n P(s_i | x_i; q^0) \log P(x_t, s_t : q) \right] \right] \\
 &= \mathop{\text{E}}_{s_t} \left[P(s_t | x_t; q^0) \log P(x_t, s_t : q) \right]
 \end{aligned}$$

Q関数の特性

・ \dot{a}_{s_t} を $\dot{a}_{s_t=w_1}^{w_c}$ とおき変える

$$Q_t(q^0, q)$$

$$= \dot{a}_{i=1}^c P(w_i | x_t; q^0) \log P(x_t, w_i : q)$$

$$= \dot{a}_{i=1}^c P(w_i | x_t; q^0) \log P(w_t) + \dot{a}_{i=1}^c P(w_i | x_t; q^0) \log P(x_i | w_t)$$

Q関数の特性

$$\begin{aligned} Q(q^0, q) &= \prod_{t=1}^n Q_t(q^0, q) \\ &= \prod_{i=1}^c \prod_{t=1}^n P(w_i | x_t; q^0) \log P(w_i) \\ &\quad + \prod_{i=1}^c \prod_{t=1}^n P(w_i | x_t; q^0) \log P(x_t | w_i) \\ &= R_1 + R_2 \end{aligned}$$

Q関数の最大化

サイコロ投げの問題の場合

[1] R_1 の最大化

$$P(w_i) = \rho_i$$

$$R_1 = \mathop{\text{ã}}_{i=1}^c \log \rho_i \mathop{\text{ã}}_{t=1}^n P(w_i | x_t; q^0) = q_i \log \rho_i$$



$\mathop{\text{ã}}_{i=1}^c \rho_i = 1$ の条件の下で最大にする ρ_i を推定値 $\hat{\rho}_i$

Q関数の最大化

$$\begin{aligned}\hat{\rho}_i &= \frac{q_i}{\prod_{j=1}^c q_j} = \frac{\prod_{t=1}^n P(w_i | x_t; q^0)}{\prod_{j=1}^c \prod_{t=1}^n P(w_j | x_t; q^0)} = \frac{\prod_{t=1}^n P(w_i | x_t; q^0)}{\prod_{t=1}^n 1} \\ &= \frac{1}{n} \prod_{t=1}^n P(w_i | x_t; q^0)\end{aligned}$$

Q関数の最大化

[2] R_2 の最大化

ごとに独立に考え下式を最大にする

$$\mathop{\text{arg}}_{t=1}^n P(w_i | x_t; q^0) \log P(x_t | w_i) \quad (i = 1, 2, \dots, c)$$

$$= \mathop{\text{arg}}_{k=1}^m \mathop{\text{arg}}_{t=1}^n d(x_t, v_k) \times P(w_i | x_t; q^0) \log P(x_t | w_i)$$

$$= \mathop{\text{arg}}_{k=1}^m r_k P(w_i | v_k; q^0) \log q_{ik}$$

Q関数の最大化

最大にする q_{ik} を推定値 \hat{q}_{ik}

$$\hat{q}_{ik} = \frac{r_k P(w_i | v_k; q^0)}{\sum_{l=1}^m r_l P(w_i | v_l; q^0)}$$

$$(k = 1, 2, \dots, m) (i = 1, 2, \dots, c)$$

EMアルゴリズムと補助関数法

補助関数法

最大化すべき目的関数 $J(\theta)$ に対し、
以下の条件を満たす補助関数 $G(\theta, \varphi)$ を導入。

(条件1) 任意の φ に対して次式が成り立つ。

$$J(q) \leq G(q, f)$$

(条件2) パラメータを $\theta = \theta^0$ に固定したとき、次式が成り立つ。

$$J(q^0) = \max_f G(q^0, f)$$

EMアルゴリズムと補助関数法

次のStep1～3を繰り返すことにより、パラメータを更新

$$\text{Step1} \quad f' \leftarrow \operatorname{argmax}_f G(q^0, f)$$

$$\text{Step2} \quad q' \leftarrow \operatorname{argmax}_q G(q, f')$$

$$\text{Step3} \quad q^0 \leftarrow q'$$

θ と ϕ の更新を繰り返すことで、 $J(\theta)$ の極大値が求まる。

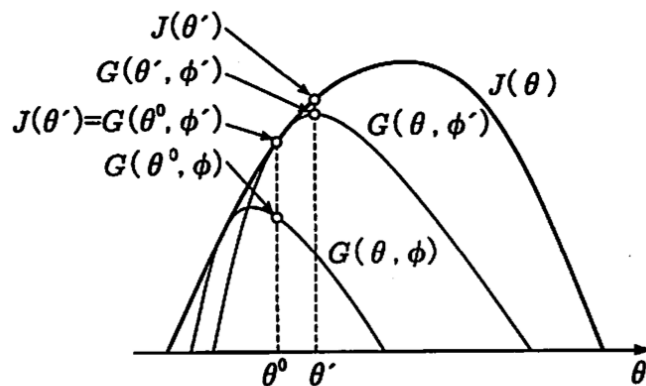


図 6.2 補助関数法による目的関数の最大化

EMアルゴリズムと補助関数法

- ・EMアルゴリズムと補助関数法は等価。
- ・補助関数法はやや遠回りにEMアルゴリズムと結果的には同じ処理を行う。